THE STATIONARY AUTOREGRESSIVE MODEL

TECHNICAL REPORT NO. 17

T. W. ANDERSON

JULY 1986

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

THE STATIONARY AUTOREGRESSIVE MODEL

T. W. Anderson

Stanford University

TECHNICAL REPORT NO. 17

July 1986

Department of Statistics

Stanford University

Stanford, California

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE
OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT
OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY
OTHER DOCUMENTATION.

1. Introduction.

There are several ways of developing the autoregressive stochastic proces as a finite–parameter model for time series analysis. In this paper we obtain the properties of the autoregressive process from a stationary stochastic process that satisfies the simple condition that a linear combination of current and past elements of the process is independent of (or alternatively uncorrelated with) all earlier elements of the process. This approach provides a coherent, clear, and rigorous exposition of the autoregressive model. The stationarity and independence imply that the roots of the associated polynomial equation are less than 1 in absolute value. The existence of the moving average representation is deduced and its form for distinct roots. The Yule–Walker equations, which are derived, determine the autocovariance sequence. Another set of parameters consists of the variance of the process and the partial autocorrelation sequence.

2. Definition of the Autoregressive Process.

A stochastic process $\{y_t\}$, which consists of a doubly infinite sequence of random variables y_t, $t = \ldots, -1, 0, 1, \ldots$, can be defined by means of the distributions of all finite sets of these variables, $y_{t_1}, \ldots, y_{t_n}$, $n = 1, 2, \ldots$. The stochastic process is *stationary* if the distribution of $y_{t_1+h}, \ldots, y_{t_n+h}$ is the same as the distribution of $y_{t_1}, \ldots, y_{t_n}$ for $h = \ldots, -1, 0, 1, \ldots$ and $n = 1, 2, \ldots$. If $\mathcal{E} y_t^2 < \infty$, then $\mathcal{E} y_t$ does not depend on t and we write $\mathcal{E} y_t = \mu$ and $\mathcal{E}(y_t - \mu)(y_{t+h} - \mu)$ does not depend on t and we write

$$(1) \qquad \mathcal{E}(y_t - \mu)(y_{t+h} - \mu) = \sigma(h), \qquad h = \ldots, -1, 0, 1, \ldots.$$

Note that $\sigma(-h) = \sigma(h)$. The quantity $\sigma(h)$ is known as the *autocovariance of order h*.

We now consider a stationary stochastic process $\{y_t\}$ with mean μ such that for some constants $\beta_0 = 1, \beta_1, \ldots, \beta_p$

$$(2) \qquad \sum_{j=0}^{p} \beta_j (y_{t-j} - \mu) = u_t$$

is independent of $y_{t-1}, y_{t-2}, \ldots$ for $t = \ldots, -1, 0, 1, \ldots$. Then $\{u_t\}$ is a stationary stochastic process because the distribution of $u_{t_1+s}, \ldots, u_{t_n+s}$ does not depend on s (since the distribution of $y_{t_1+s}, y_{t_1+s-1}, \ldots, y_{t_1+s-p}, \ldots, y_{t_n+s}, \ldots, y_{t_n+s-p}$ does not depend on s). Furthermore, u_t is independent of $u_{t-1} = \sum_{j=0}^{p} \beta_j (y_{t-j-1} - \mu), u_{t-2} = \sum_{j=0}^{p} \beta_j (y_{t-j-2} - \mu), \ldots$. Thus the u_t's are independently and identically distributed. Since $\mathcal{E} y_t = \mu$, then $\mathcal{E} u_t = 0$. If $\mathcal{E} y_t^2 < \infty$, we can write $\mathcal{E} u_t^2 = \sigma^2$, which we assume to be positive. (If $\sigma^2 = 0$, the process satisfies a linear relation with probability 1 and is deterministic.) The process $\{y_t\}$ is called an *autoregressive process of order* p, and (2) is called a *stochastic difference equation* of degree p.

Now multiply the left–hand and right–hand sides of the stochastic difference equation (2) by the left–hand and right–hand sides of $y_t - \mu = u_t - \sum_{j=1}^{p} \beta_j (y_{t-j} - \mu)$, respectively, and take expected values, assuming $\mathcal{E}(y_t - \mu)^2 < \infty$. Since u_t is independent of $y_{t-1}, \ldots, y_{t-p}$, we obtain

$$(3) \qquad \sum_{j=0}^{p} \beta_j \sigma(j) = \sigma^2.$$

When we multiply both sides of (2) by $y_{t-s} - \mu$ and take expected values, we obtain

$$(4) \qquad \sum_{j=0}^{p} \beta_j \sigma(s-j) = 0, \qquad s = 1, 2, \ldots.$$

Equations (3) and (4) will be called the *Yule–Walker equations*. (Many authors do not include (3) in "the Yule-Walker equations.")

Now we shall show that the roots of the "associated polynomial equation are less than 1 in absolute value. If $p = 1$, (3) and (4) for $s = 1$ are

$$(5) \qquad \sigma(0) + \beta_1 \sigma(1) = \sigma^2,$$

2

(6)
$$\sigma(1) + \beta_1 \sigma(0) = 0.$$

These yield

(7)
$$(1 - \beta_1^2)\sigma(0) = \sigma^2.$$

Since $\sigma^2 > 0$, it follows that $\sigma(0) > 0$ and $(1 - \beta_1^2) > 0$; that is $|\beta_1| < 1$. (If $\sigma^2 = 0$ and $\sigma(0) > 0$, then either $\beta_1 = 1, \sigma(1) = -\sigma(0)$, and $y_t = -y_{t-1}$ with probability 1 or $\beta_1 = -1, \sigma(1) = \sigma(0)$, and $y_t = y_{t-1}$ with probability 1.)

Now suppose $p = 2$ and the roots of

(8)
$$x^2 + \beta_1 x + \beta_2 = 0$$

are complex conjugate $(\beta_1^2 - 4\beta_2 < 0)$; denote the roots as $\alpha e^{\pm i\theta}$, where $0 < \alpha$ and $0 < \theta < \pi$. Then (3) and (4) for $s = 1$ and 2 are

(9)
$$\sigma(0) - 2\alpha \cos \theta \sigma(1) + \alpha^2 \sigma(2) = \sigma^2,$$

(10)
$$-2\alpha \cos \theta \sigma(0) + (1 + \alpha^2)\sigma(1) = 0,$$

(11)
$$\sigma(2) - 2\alpha \cos \theta \sigma(1) + \alpha^2 \sigma(0) = 0.$$

Subtraction of (11) from (9) yields

(12)
$$(1 - \alpha^2)[\sigma(0) - \sigma(2)] = \sigma^2.$$

3

Since $\sigma^2 > 0$, we see that $\alpha < 1$ and $\sigma(2) < \sigma(0)$ because $\sigma(2) \leq \sigma(0)$ by the Cauchy–Schwarz inequality. (If $\sigma^2 = 0, \sigma(0) > 0$, and the roots are not real, $\alpha = 1$.)

Now we consider p in general. Let $x_1, \ldots, x_p$ be the roots of the *associated polynomial equation*

$$(13) \qquad \sum_{j=0}^{p} \beta_j x^{p-j} = 0.$$

Then the associated polynomial can be written

$$(14) \qquad \sum_{j=0}^{p} \beta_j x^{p-j} = \prod_{i=1}^{p} (x - x_i),$$

and the stochastic difference equation can be written

$$(15) \qquad u_t = \sum_{j=0}^{p} \beta_j \mathcal{L}^j (y_t - \mu) = \prod_{i=1}^{p} (1 - x_i \mathcal{L})(y_t - \mu),$$

where $\mathcal{L}^j y_t = y_{t-j}$. If x_1 is a real root, we can write the stochastic difference equation as

$$(16) \qquad w_t - x_1 w_{t-1} = u_t,$$

where

$$(17) \qquad w_t = \prod_{i=2}^{p} (1 - x_i \mathcal{L})(y_t - \mu).$$

Then u_t is independent of $w_{t-1}, w_{t-2}, \ldots$, and $\{w_t\}$ is a stationary autoregressive process of order 1 and hence $|x_1| < 1$. If x_1 and x_2 are conjugate complex roots, $x_1 = \alpha e^{i\theta}$ and $x_2 = \alpha e^{-i\theta}$, then

4

$$(18) \qquad v_t - 2\alpha \cos\theta v_{t-1} + \alpha^2 v_{t-2} = u_t,$$

where

$$(19) \qquad v_t = \prod_{i=3}^{p}(1 - x_i \mathcal{L})(y_t - \mu).$$

Then u_t is independent of $v_{t-1}, v_{t-2}, \ldots$, and $\{v_t\}$ is a stationary autoregressive process of order 2. From the above treatment of the case $p = 2$, we know that $|x_1| = |x_2| = \alpha < 1$. We summarize the above results in the following theorem.

Theorem 1. Let $\{y_t\}$ be a stationary stochastic process with mean $\mathcal{E} y_t = \mu$ and finite variance such that (2) holds for suitable $\beta_1, \ldots, \beta_p$ ($\beta_0 = 1$) and $\mathcal{E} u_t^2 = \sigma^2 > 0$ with u_t independent of $y_{t-1}, y_{t-2}, \ldots$. Then u_t is a sequence of independently and identically distributed random variables. The second–order moments $\mathcal{E}(y_t - \mu)(y_{t+h} - \mu) = \sigma(h)$ satisfy (3) and (4). The roots of the associated polynomial equation (13) are less than 1 in absolute value.

3. The Moving Average Representation.

Now we want to show that y_t can be expressed as an infinite linear combination of the independent u_t. In the case of $p = 1$ we write

$$(20) \qquad y_t = \phi y_{t-1} + u_t.$$

If we replace y_{t-1} in (20) by $\phi y_{t-2} + u_{t-1}$ (which is (20) with t replaced by $t - 1$), we obtain

$$(21) \qquad y_t = u_t + \phi u_{t-1} + \phi^2 y_{t-2}.$$

If we substitute successively, we obtain

5

(22)
$$y_t = u_t + \phi u_{t-1} + \ldots + \phi^s u_{t-s} + \phi^{s+1} y_{t-(s+1)}.$$

From (22) we obtain

(23)
$$\mathcal{E}\left[y_t - (u_t + \phi u_{t-1} + \ldots + \phi^s u_{t-s})\right]^2 = \phi^{2(s+1)} \mathcal{E} y_{t-(s+1)}^2.$$

Since $|\phi| = |\beta_1| < 1$, the right–hand side of (23) goes to 0 as s increases. Then we write

(24)
$$y_t = \sum_{r=0}^{\infty} \phi^r u_{t-r}$$

and say that the series on the right *converges in the mean* (of in the mean square or in quadratic mean) to y_t.

Now let us turn to the general case. The lag operator $\mathcal{L}$ is a linear operator for which polynomials are defined. The autoregressive model (2) with $\mu = 0$ can be written

(25)
$$\left(\sum_{j=0}^{p} \beta_j \mathcal{L}^j \right) y_t = u_t.$$

Formally (25) can be inverted to give

(26)
$$y_t = \left(\sum_{j=0}^{p} \beta_j \mathcal{L}^j \right)^{-1} u_t$$
$$= \left(\sum_{s=0}^{\infty} \delta_s \mathcal{L}^s \right) u_t$$
$$= \sum_{s=0}^{\infty} \delta_s u_{t-s},$$

where the coefficients $\delta_0, \delta_1, \ldots$ are defined by the identity

$$(27) \qquad 1 = \sum_{j=0}^{p} \beta_j z^j \sum_{s=0}^{\infty} \delta_s z^s$$

$$= \sum_{j=0}^{p} \sum_{s=0}^{\infty} \beta_j \delta_s z^{j+s}$$

$$= \sum_{r=0}^{\infty} \sum_{j=0}^{\min(r,p)} \beta_j \delta_{r-j} z^r,$$

where $r = j + s$ and the limits of the last sum on j depend on the value of r. The coefficient of each power of z on the left–hand side of (27) is equal to the coefficient of that power on the right–hand side. Thus

$$1 = \beta_0 \delta_0 = \delta_0,$$

$$(28) \qquad 0 = \beta_0 \delta_1 + \beta_1 \delta_0 = \delta_1 + \beta_1,$$

$$\vdots$$

$$0 = \beta_0 \delta_{p-1} + \beta_1 \delta_{p-2} + \ldots + \beta_{p-1} \delta_0,$$

$$(29) \qquad 0 = \beta_0 \delta_t + \beta_1 \delta_{t-1} + \ldots + \beta_p \delta_{t-p}, \qquad t = p, p+1, \ldots.$$

We look for a sequence $\delta_0, \delta_1, \ldots$ that satisfies the p equations (28) and the set of equations (29). Equations (28) determine $\delta_0, \ldots, \delta_{p-1}$ successively and (29) determines $\delta_p, \delta_{p+1}, \ldots$ in turn; the solution to these equations exists and is unique.

Does (26) converge in the mean? We note that

$$(30) \qquad (1 - x_i z)^{-1} = \sum_{s=0}^{\infty} x_i^s z^s$$

converges uniformly for $|z| \leq \gamma/|x_i|$ for $0 < \gamma < 1$. Hence,

7

$$(31) \qquad \sum_{s=0}^{\infty} \delta_s z^s = (1 + \beta_1 z + \ldots + \beta_p z^p)^{-1} = \prod_{i=1}^{p} \left(\sum_{s=0}^{\infty} x_i^s z^s \right)$$

converges uniformly for $|z| \leq \gamma / \max_{i=1,\ldots,p} |x_i|$, which is greater than 1 if γ is large enough. Hence, $\delta_s \to 0$ as $s \to \infty$. We write for $m > p$

$$(32) \qquad \begin{aligned} \sum_{s=0}^{m} \delta_s u_{t-s} &= \sum_{s=0}^{m} \delta_s \sum_{j=0}^{p} \beta_j y_{t-s-j} \\ &= \sum_{s=0}^{m} \sum_{j=0}^{p} \beta_j \delta_s y_{t-s-j} \\ &= \sum_{r=0}^{m+p} \sum_{j=\max(0,r-m)}^{\min(r,p)} \beta_j \delta_{r-j} y_{t-r} \\ &= y_t + \sum_{r=m+1}^{m+p} \sum_{j=r-m}^{p} \beta_j \delta_{r-j} y_{t-r} \end{aligned}$$

by comparison with (27). From this we obtain for $m > p$

$$(33) \qquad \mathcal{E} \left(y_t - \sum_{s=0}^{m} \delta_s u_{t-s} \right)^2 = \mathcal{E} \left(\sum_{r=m+1}^{m+p} \sum_{j} \beta_j \delta_{r-j} y_{t-r} \right)^2 .$$

The right–hand side is a quadratic form in $\delta_{m+1-p}, \ldots, \delta_{m+p}$. Since $\delta_s \to 0$ as $s \to \infty$, the right–hand side converges to 0 and (26) converges in the mean as $m \to \infty$.

Theorem 2. If y_t, $t = \ldots, -1, 0, 1, \ldots$, satisfies (2) with u_t independent of $y_{t-1}, y_{t-2}, \ldots$ and $\mathcal{E} y_t = \mu$, then

$$(34) \qquad y_t = \mu + \sum_{s=0}^{\infty} \delta_s u_{t-s}$$

converges in the mean, where $\delta_0 = 1, \delta_1, \ldots$ are determined by (28) and (29).

The expression (34) is called the *moving average representation* of the autoregressive process $\{y_t\}$.

8

We shall now find an explicit expression for δ_s when the roots of the associated polynomial equation are distinct.

The equation (29) is a *homogeneous difference equation*. (It is homogeneous of degree 1 in the δ_t's, and it is a difference equation because it can be written as a polynomial in $\nabla = 1 - \mathcal{L}$ operating on δ_t.) The case $p = 1$ suggests that a solution to the homogeneous difference equation (29) may be of the form $\delta_t = x^t$. Then the right–hand side of

$$(35) \qquad \sum_{j=0}^{p} \beta_j x^{t-j} = x^{t-p} \sum_{j=0}^{p} \beta_j x^{p-j}$$

is 0 if x satisfies the associated polynomial equation (13); that is, if x is one of the roots $x_1, \ldots, x_p$ of (13). It follows that

$$(36) \qquad \delta_t = \sum_{i=1}^{p} k_i x_i^t, \qquad t = 0, 1, \ldots,$$

is also a solution. If we substitute (36) into (28), we obtain

$$1 = \sum_{i=1}^{p} k_i,$$

$$(37) \qquad 0 = \sum_{i=1}^{p} k_i x_i + \beta_1 \sum_{i=1}^{p} k_i = \sum_{i=1}^{p} (x_i + \beta_1) k_i,$$

$$\vdots$$

$$0 = \sum_{i=1}^{p} k_i x_i^{p-1} + \beta_1 \sum_{i=1}^{p} k_i x_i^{p-2} + \ldots + \beta_{p-1} \sum_{i=1}^{p} k_i = \sum_{i=1}^{p} (x_i^{p-1} + \beta_1 x_i^{p-2} + \ldots + \beta_{p-1}) k_i.$$

The detrminant of the coefficients of $k_1, \ldots, k_p$ in (37) is

$$(38)$$

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 + \beta_1 & x_2 + \beta_1 & \cdots & x_p + \beta_1 \\ \vdots & \vdots & & \vdots \\ x_1^{p-1} + \beta_1 x_1^{p-2} + \ldots + \beta_{p-1} & x_2^{p-1} + \beta_1 x_2^{p-2} + \ldots + \beta_{p-1} & \cdots & x_p^{p-1} + \beta_1 x_p^{p-2} + \ldots + \beta_{p-1} \end{vmatrix}$$

9

$$
= \begin{vmatrix} 1 & 0 & \cdots & 0 \\ \beta_1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \beta_{p-1} & \beta_{p-2} & \cdots & 1 \end{vmatrix} \circ \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_p \\ \vdots & \vdots & & \vdots \\ x_1^{p-1} & x_2^{p-1} & \cdots & x_p^{p-1} \end{vmatrix}
$$

$$
= \prod_{j < i} (x_i - x_j).
$$

This (Vandermonde) determinant is different from 0 if and only if the roots are distinct. Hence, there is a unique solution to (28) of the form (36) if and only if the roots of (13) are distinct. Then $\delta_t = \sum_{i=1}^{p} k_i x_i^t$ is a solution to (28) and (29) and is the only solution to (28). Since δ_t is uniquely determined by $\delta_{t-1}, \ldots, \delta_{t-p}$ in (29), $t = p, p+1, \ldots$, the solution of $\delta_t = \sum_{i=1}^{p} k_i x_i^t$ is the unique solution to (28) and (29) if the roots of (13) are distinct.

Theorem 3. The unique solution to the homogeneous difference equation (29) satisfying the conditions (28) is $\delta_t = \sum_{i=1}^{p} k_i x_i^t$, where $k_1, \ldots, k_p$ satisfy (37) and $x_1, \ldots, x_p$ are the roots of (13), if $x_1, \ldots, x_p$ are distinct.

If $p = 1$, $\delta_r = (-\beta_1)^r$, an exponential function of r. If $p = 2$ and x_1 and x_2 are different, $k_1 = x_1/(x_1 - x_2)$, $k_2 = -x_2/(x_1 - x_2)$, and

$$
(39) \qquad \delta_r = \frac{x_1^{r+1} - x_2^{r+1}}{x_1 - x_2}, \qquad r = 0, 1, \ldots.
$$

If x_1 and x_2 are real, δ_r is a linear combination of two exponential functions of r. If the roots are complex, we may write them $x_1 = \alpha e^{i\theta}$ and $x_2 = \alpha e^{-i\theta}$. Then

$$
(40) \qquad \delta_r = \alpha^r \frac{e^{i\theta(r+1)} - e^{i\theta(r+1)}}{e^{i\theta} - e^{-i\theta}}
$$

$$
= \alpha^r \frac{\sin \theta (r + 1)}{\sin \theta},
$$

a damped sine function of r.

4. The Autocovariance Sequence.

Now we find an explicit expression for $\{\sigma(h)\}$ when the roots of the associated polynomial equation are distinct. To do this we first demonstrate that (3) and the first p equations in (4) determine $\sigma(0), \sigma(1), \ldots, \sigma(p)$ uniquely regardless of whether the roots are distinct. Since $\sigma(h) = \sigma(-h), h = 1, \ldots, p-1$, these $p+1$ equations can be written

$$(41) \qquad \sigma(0) + \beta_1\sigma(1) + \ldots + \beta_p\sigma(p) = \sigma^2,$$

$$\beta_1\sigma(0) + (1+\beta_2)\sigma(1) + \ldots + \beta_p\sigma(p-1) = 0,$$

$$\vdots$$

$$\beta_p\sigma(0) + \beta_{p-1}\sigma(1) + \ldots + \sigma(p) = 0.$$

The determinant of the matrix of coefficients is

$$(42) \qquad \begin{vmatrix} 1 & \beta_1 & \beta_2 & \ldots & \beta_{p-1} & \beta_p \\ \beta_1 & 1+\beta_2 & \beta_3 & \ldots & \beta_p & 0 \\ \beta_2 & \beta_1+\beta_3 & 1+\beta_4 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \beta_{p-1} & \beta_{p-2}+\beta_p & \beta_{p-3} & \ldots & 1 & 0 \\ \beta_p & \beta_{p-1} & \beta_{p-2} & \ldots & \beta_1 & 1 \end{vmatrix}.$$

If we multiply the $(j+1)st$ column by $x_1^j + x_1^{-j}$ and add the result to 2 times the first column, $j = 1, \ldots, p$, we obtain as the first column the transpose of

$$(43) \qquad \left(\sum_{j=0}^{p}\beta_j x_1^j + \sum_{j=0}^{p}\beta_j x_1^{-j}, x_1^{-1}\sum_{j=0}^{p}\beta_j x_1^j + x_1\sum_{j=0}^{p}\beta_j x_1^{-j}, \ldots, x_1^{-p}\sum_{j=0}^{p}\beta_j x_1^j + x_1^p\sum_{j=0}^{p}\beta_j x_1^{-j} \right)$$

$$= \sum_{j=0}^{p}\beta_j x_1^j(1, x_1, \ldots, x_1^p)$$

11

because $\sum_{j=0}^{p} \beta_j x_1^{-j} = 0$. We assume that $\beta_p \neq 0$ and hence $x_j \neq 0$, $j = 1, \ldots, p$. Thus

$$(44) \qquad \sum_{j=0}^{p} \beta_j x_1^j = \prod_{i=1}^{p} (1 - x_i x_1)$$

is a factor of the determinant. Since this operation can be done with any root,

$$(45) \qquad \prod_{i \leq j} (1 - x_i x_j)$$

is a factor. The term in the determinant of highest degree in $x_1, \ldots, x_p$ is

$$(46) \qquad \pm \beta_p^{p+1} = \pm \prod_{i=1}^{p} x_i^{p+1},$$

which is the degree of (45). Since the determinant is 1 for $x_1 = \ldots = x_p = 0$, the determinant is (45). This is different from 0 because $|x_i| < 1$, $i = 1, \ldots, p$. Thus the Yule–Walker equations determine $\sigma(0), \ldots, \sigma(p)$ uniquely as the solution to (41).

The equation (4) is a homogeneous difference equation for $\sigma(h)$, $h = -(p-1), -(p-2), \ldots$, and $\sigma(h) = \sum_{i=1}^{p} c_i x_i^h$ satisfies it for any $c_1, \ldots, c_p$. However, if the roots are distinct there is only one set of $c_1, \ldots, c_p$ that satisfies $\sigma(h) = \sum_{i=1}^{p} c_i x_i^h$ for $h = -(p-1), \ldots, 0$; that is,

$$(47) \qquad \begin{bmatrix} \sigma(0) \\ \sigma(-1) \\ \vdots \\ \sigma(-p+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1^{-1} & x_2^{-1} & \cdots & x_p^{-1} \\ \vdots & \vdots & & \vdots \\ x_1^{-(p-1)} & x_2^{-(p-1)} & \cdots & x_p^{-(p-1)} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}.$$

The determinant of the matrix of coefficients of $c_1, \ldots, c_p$ in (47) in the Vandermonde determinant $\prod_{i>j} (x_i^{-1} - x_j^{-1})$, which is nonzero.

Theorem 4. If the roots of the associated polynomial equation are distinct and $\beta_p \neq 0$, the autocovariances are given by $\sigma(h) = \sum_{i=1}^{p} c_i x_i^h$, $h = 1-p, 2-p, \ldots$, where $c_1, \ldots, c_p$ satisfy (47) or equivalently (3) and $\sigma(h) = \sigma(-h)$, $h = 1, \ldots, p-1$.

If $p = 1, \sigma(h) = (-\beta_1)^h \sigma^2 / (1 - \beta_1^2)$, $h = 0, 1, \ldots$, an exponential function of h. If $p = 2$ and x_1 and x_2 are different,

$$\text{(48)} \qquad \sigma(h) = \frac{\sigma^2}{(x_1 - x_2)(1 - x_1 x_2)} \left(\frac{x_1^{h+1}}{1 - x_1^2} - \frac{x_2^{h+1}}{1 - x_2^2} \right).$$

If x_1 and x_2 are real, $\sigma(h)$ is a linear combination of two exponential functions. If $p = 2$ and x_1 and x_2 are conjugate complex, $\alpha e^{\pm i\theta}$, then (48) can be written

$$\text{(49)} \qquad \sigma(h) = \frac{\sigma^2 \alpha^h \left[\sin\theta(h+1) - \alpha^2 \sin\theta(h-1) \right]}{(1 - \alpha^2) \sin\theta \left[1 - 2\alpha^2 \cos 2\theta + \alpha^4 \right]},$$

$$= \frac{\sigma^2 \alpha^h h \cos(\theta h - \psi)}{(1 - \alpha^2) \sin\theta \sqrt{1 - 2\alpha^2 \cos 2\theta + \alpha^4}},$$

where

$$\text{(50)} \qquad \tan\psi = \frac{(1 - \alpha^2)\cos\theta}{(1 + \alpha^2)\sin\theta}.$$

Here we have assumed that $\{y_t\}$ is strictly stationary and that u_t defined by (2) is indepenent of $y_{t-1}, y_{t-2}, \ldots$. Then the u_t's are independently identically distributed. A similar development can be carried out on the assumptions that $\{y_t\}$ is stationary in the wide sense and that u_t is uncorrelated with $y_{t-1}, y_{t-2}, \ldots$. Then the u_t's are uncorrelated and have the same variance.

An alternative approach to the stationary autoregressive model is to assume $\{u_t\}$ consists of independently identically distributed random variables with $\mathcal{E} u_t = 0$ and $\mathcal{E} u_t^2 = \sigma^2$. Then define y_t by (34), where $\{\delta_s\}$ satisfies (28) and (29) for some $\beta_1, \ldots, \beta_p$ with the roots of (13) less than 1 in absolute value. Then $\{y_t\}$ satisfies (2) and u_t is independent of $y_{t-1}, y_{t-2}, \ldots$.

5. The Direction of Time.

What gives time the direction is that u_t is independent of the observable past: $y_{t-1}, y_{t-2}, \ldots$. If $\{y_t\}$ is Gaussian (that is, all finite distributions are normal), then $\{y_t\}$ is determined by $\mathcal{E} y_t = \mu$ and $\mathcal{E}(y_t - \mu)(y_{t+h} - \mu) = \sigma(h), h = 0, 1, \ldots$. Since $\sigma(h) = \sigma(-h)$,

the covariance function does not distinguish between the sequence $t = \ldots, -1, 0, 1, \ldots$ and the sequence $t = \ldots, 1, 0, -1, \ldots$. In fact, for $\beta_1, \ldots, \beta_p$ and σ^2 given, the random variable

$$(51) \qquad \sum_{j=0}^{p} \beta_j y_{t+h} = w_t$$

is independent of $y_{t+1}, y_{t+2}, \ldots$, $\{w_t\}$ is a sequence of independently identically distributed random variables, and the Yule–Walker equations hold. For example, if $\sigma(h) = (-\beta_1)^{|h|} \sigma(0)$, $h = \pm 1, \pm 2, \ldots$, then (51) is

$$(52) \qquad y_t + \beta_1 y_{t+1} = w_t,$$

and $y_t = \sum_{s=0}^{\infty} (-\beta_1)^s w_{t+s}$.

Let us look briefly at the case in which some roots of the polynomial equation are larger than 1 in absolute value. Suppose $|x_i| > 1$, $i = 1, \ldots, q$, $|x_i| < 1$, $i = q + 1, \ldots, p$. We write the stochastic difference equation as

$$(53) \qquad \begin{aligned} u_t &= \sum_{r=0}^{p} \beta_r \mathcal{L}^r y_t \\ &= \prod_{i=1}^{p} (1 - x_i \mathcal{L}) y_t. \end{aligned}$$

The inverse of (53) is

$$(54) \qquad \begin{aligned} y_t &= \prod_{i=1}^{p} (1 - x_i \mathcal{L})^{-1} u_t \\ &= \prod_{i=1}^{q} (1 - x_i \mathcal{L})^{-1} \prod_{i=q+1}^{p} (1 - x_i \mathcal{L})^{-1} u_t \\ &= \prod_{i=1}^{q} \left(-\frac{1}{x_i} \right) \mathcal{P} \left(1 - \frac{1}{x_i} \mathcal{P} \right)^{-1} \prod_{i=q+1}^{p} (1 - x_i \mathcal{L})^{-1} u_t \end{aligned}$$

where $\mathcal{P} = \mathcal{L}^{-1}$. Each term $(1 - x_i^{-1} \mathcal{P})^{-1}$ can be expanded in a power series in $\mathcal{P}$, and each term $(1 - x_i \mathcal{L})^{-1}$ can be expanded in a power series in $\mathcal{L}$. If $1 \leq q < p$, the operator will be a power series in $\mathcal{L}$ ($= \mathcal{P}^{-1}$) and in $\mathcal{P}$ ($= \mathcal{L}^{-1}$); the indicated infinite series in $\ldots, u_{t-1}, u_t, u_{t+1}, \ldots$ will be doubly infinite. If $q = p$ (all roots in absolute value greater than 1), the operator is

14

an infinite power series in p (only nonnegative powers); the indicated infinite series in the random variables involves $u_{t+p}, u_{t+p+1}, \ldots$. This discussion is in purely formal terms, but it can be justified in the manner used for the case of all roots less than 1 in absolute value. In fact, if the linear form $\prod_{i=1}^{p}(1 - x_i \mathcal{L})y_t$ is replaced by the linear form (for any x_i's)

$$(55) \qquad \prod_{i=1}^{r}(p - x_i^{-1}) \prod_{i=r+1}^{p}(1 - x_i \mathcal{L})y_t = u_t^*$$

the residuals u_t^* are also uncorrelated (though not necessarily independent).

Now consider the special case of one root equal to 1

$$(56) \qquad y_t = y_{t-1} + u_t,$$

and assume $\mathcal{E} u_t = 0$, $\mathcal{E} u_t^2 = \sigma^2$. Then

$$(57) \qquad y_t - y_{t-s} = u_t + u_{t-1} + \ldots + u_{t-s+1},$$

and

$$(58) \qquad \mathcal{E}(y_t - y_{t-s})^2 = \mathcal{E} y_t^2 + \mathcal{E} y_{t-s}^2 - 2\mathcal{E} y_t y_{t-s} = s\sigma^2.$$

If the process is stationary, $\mathcal{E} y_t^2 = \mathcal{E} y_{t-s}^2$; from (58) we obtain

$$(59) \qquad \mathcal{E} y_t y_{t-s} = \mathcal{E} y_t^2 - \frac{1}{2}s\sigma^2, \qquad s = 1, 2, \ldots.$$

This can hold for all $s > 0$ only if $\sigma^2 = 0$ and then $y_t = y_{t-s}$ with probability 1. Intuitively, we see that unless the variance of u_t is 0, (56) implies that the variance of y_t increases with t, but this fact is contrary to stationarity.

A more general case is

$$(60) \qquad (1 - \mathcal{L})\prod_{i=2}^{p}(1 - x_i \mathcal{L})y_t = u_t,$$

where $|x_i| \neq 1, i = 2, \ldots, p$. If $\prod_{i=2}^{p}(1 - x_i \mathcal{L})y_t = z_t$, then $(1 - \mathcal{L})z_t = u_t$ and $z_t = z_{t-s}$ with probability 1, say $z_t = z$. Thus

$$(61) \qquad \prod_{i=2}^{p}(1 - x_i \mathcal{L})y_t = z,$$

and $y_t = \sum_{s=-\infty}^{\infty} \delta_s z$; that is, $y_t = y_{t-s}$ with probability 1. (z can be a random variable.)

6. Fluctuations of the Time Series.

The typical time series generated by an autoregression model fluctuates up and down. Its oscillations are not regular, but tend to have an average length which depends on the difference equation. If we think of (2) as generating the series for successive values of t we see that each set of p y_t's directly affects the next y_t. If $p = 2$, we can write the equation

$$(62) \qquad \begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t \\ &= (\phi_1 + \phi_2)y_{t-1} - \phi_2(y_{t-1} - y_{t-2}) + u_t, \end{aligned}$$

which indicates that the direct effect of preceding y_r's on y_t involves the value of y_{t-1} and the change between y_{t-2} and y_{t-1}. This effect will generally have a tendency to produce fluctuations.

Another way of looking at the process is in terms of the representation of $y_t = \sum_{r=0}^{\infty} \delta_r u_{t-r}$. A given u_s will affect a subsequent y_{s+q} according to the coefficient δ_q. Since these coefficients oscillate, the effect on successive y_r's fluctuates, tending to produce fluctuations in the series y_t.

The fluctuating behavior is also indicated by the covariance function. Since $\sigma(s) \neq 0$ (usually), there is some statistical association between y_t and y_{t+s}, though it tends to decrease as s (> 0) increases. This association (measured by the correlation coefficient) tends to fluctuate. For example, if there is a pair of complex roots, the trigonometric functions oscillate and the association may increase with s in some intervals (of s).

7. Partial Autocorrelations.

A stationary stochastic process $\{y_t\}$ with finite second moment defines an autocovariance sequence $\{\sigma(h)\}$. In turn, each set $\sigma(0), \sigma(1), \ldots, \sigma(p)$ defines a set of coefficients $\phi_1(p), \ldots, \phi_p(p)$ by the p Yule–Walker equations

$$(63) \qquad \sum_{j=1}^{p} \sigma(s-j)\phi_j(p) = \sigma(s), \qquad s = 1, \ldots, p.$$

We shall now show that the autoregressive coefficient $\phi_p(p)$ in the fitted AR(p) model is identical to the partial correlation coefficient between y_t and y_{t-p} "holding $y_{t-1}, \ldots, y_{t-p+1}$ fixed." This partial correlation coefficient is the correlation between the residual of y_t regressed on $y_{t-1}, \ldots, y_{t-p+1}$ and the residual of y_{t-p} regresssed on $y_{t-1}, \ldots, y_{t-p+1}$. We suppose $\mathcal{E}\, y_t = 0$.

Define

$$(64) \qquad \underset{\sim}{y}_{t-1}^{(p-1)} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \mathcal{E}\, y_t \underset{\sim}{y}_t^{(p-1)} = \underset{\sim}{\sigma}_{p-1} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(p-1) \end{bmatrix},$$

$$(65) \qquad \mathcal{E}\, \underset{\sim}{y}_{t-1}^{(p-1)} \underset{\sim}{y}_{t-1}^{(p-1)\prime} = \underset{\sim}{\Sigma}_{p-1} = \begin{bmatrix} \sigma(0) & \sigma(1) & \ldots & \sigma(p-2) \\ \sigma(1) & \sigma(0) & \ldots & \sigma(p-3) \\ \vdots & \vdots & & \vdots \\ \sigma(p-2) & \sigma(p-3) & \ldots & \sigma(0) \end{bmatrix}.$$

Then the vector of regression coefficients $\underset{\sim}{\phi}_{p-1} = \left[\phi_1(p-1), \ldots, \phi_{p-1}(p-1)\right]'$ of y_t on $\underset{\sim}{y}_{t-1}^{(p-1)} = (y_{t-1}, \ldots, y_{t-p+1})'$ is the solution to

$$(66) \qquad \underset{\sim}{\Sigma}_{p-1} \underset{\sim}{\phi}_{p-1} = \underset{\sim}{\sigma}_{p-1}.$$

The vector of regression coefficients of y_{t-p} on $y_{t-p+1}, \ldots, y_{t-1}$ is defined by the same equation. (The covariances of y_r and y_s depends only on $|r - s|$, not the sign of $r - s$.) Thus the vector of

17

regression coefficients of y_{t-p} on $\underset{\sim}{y}_{t-1}^{(p-1)}$ is $\underset{\sim}{\tilde{\phi}}_{p-1} = [\phi_{p-1}(p-1), \ldots, \phi_1(p-1)]'$. The variances of the residuals are

$$
(67) \qquad \mathcal{E}(y_t - \underset{\sim}{\phi}_{p-1}'\underset{\sim}{y}_{t-1}^{(p-1)})^2 = \sigma(0) = \underset{\sim}{\phi}_{p-1}'\underset{\sim}{\sigma}_{p-1}\underset{\sim}{\phi}_{p-1},
$$

$$
(68) \qquad \mathcal{E}(y_{t-p} - \underset{\sim}{\tilde{\phi}}_{p-1}'\underset{\sim}{y}_{t-1}^{(p-1)})^2 = \sigma(0) - \underset{\sim}{\tilde{\phi}}_{p-1}'\underset{\sim}{\Sigma}_{p-1}\underset{\sim}{\tilde{\phi}}_{p-1};
$$

these two are the same. The covariance between the two residuals is

$$
(69) \qquad \mathcal{E}(y_t - \underset{\sim}{\phi}_{p-1}'\underset{\sim}{y}_{t-1}^{(p-1)})(y_{t-p} - \underset{\sim}{\tilde{\phi}}_{p-1}\underset{\sim}{y}_{t-1}^{(p-1)}) = \sigma(p) - \underset{\sim}{\phi}_{p-1}'\underset{\sim}{\Sigma}_{p-1}\underset{\sim}{\tilde{\phi}}_{p-1}.
$$

The *partial autocorrelation* between y_t and y_{t-p} given $\underset{\sim}{y}_{t-1}^{(p-1)}$ is

$$
(70) \qquad \mathrm{PACF}(p) = \frac{\sigma(p) - \underset{\sim}{\phi}_{p-1}'\underset{\sim}{\Sigma}_{p-1}\underset{\sim}{\tilde{\phi}}_{p-1}}{\sigma(0) - \underset{\sim}{\phi}_{p-1}'\underset{\sim}{\Sigma}_{p-1}\underset{\sim}{\phi}_{p-1}}.
$$

Now let us write the equation for the vector of regression coefficients of y_t on $y_{t-1}, \ldots, y_{t-p}$ in partitioned form

$$
(71) \qquad \begin{bmatrix} \underset{\sim}{\Sigma}_{p-1} & \underset{\sim}{\tilde{\sigma}}_{p-1} \\ \underset{\sim}{\tilde{\sigma}}_{p-1}' & \sigma(0) \end{bmatrix} \begin{bmatrix} \underset{\sim}{\phi}^{(1)}(p) \\ \phi_p(p) \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\sigma}_{p-1} \\ \sigma(p) \end{bmatrix},
$$

where $\underset{\sim}{\tilde{\sigma}}_{p-1} = [\sigma(p-1), \ldots, \sigma(1)]'$. The partitioned equations are

$$
(72) \qquad \underset{\sim}{\Sigma}_{p-1}\underset{\sim}{\phi}^{(1)}(p) + \underset{\sim}{\tilde{\sigma}}_{p-1}\phi_p(p) = \underset{\sim}{\sigma}_{p-1},
$$

$$
(73) \qquad \underset{\sim}{\tilde{\sigma}}_{p-1}^{(1)'}\underset{\sim}{\phi}^{(1)}(p) + \sigma(0)\phi_p(p) = \sigma(p).
$$

The first equation yields

(74)
$$\underset{\sim}{\phi}^{(1)}(p) = \underset{\sim}{\Sigma}_{p-1}^{-1}\underset{\sim}{\sigma}_{p-1} - \underset{\sim}{\Sigma}_{p-1}^{-1}\underset{\sim}{\tilde{\sigma}}_{p-1}\phi_p(p)$$

$$= \underset{\sim}{\phi}_{p-1} - \phi_p(p)\underset{\sim}{\tilde{\phi}}_{p-1}.$$

Substitution into (73) yields

(75)
$$\left[\sigma(0) - \underset{\sim}{\tilde{\sigma}}'_{p-1}\underset{\sim}{\Sigma}_{p-1}^{-1}\underset{\sim}{\tilde{\sigma}}_{p-1}\right]\phi_p(p) = \sigma(p) - \underset{\sim}{\tilde{\sigma}}'_{p-1}\underset{\sim}{\Sigma}_{p-1}^{-1}\underset{\sim}{\sigma}_{p-1}.$$

Comparison of (70) and (75) shows that $\phi_p(p) = \text{PACF}(p)$.

These equations lead to the Levinson–Durbin recursive relations (74) and

(76)
$$\phi_p(p) = \frac{\underset{\sim}{\sigma}(p) - \underset{\sim}{\tilde{\sigma}}'_{p-1}\underset{\sim}{\phi}_{p-1}}{\underset{\sim}{\sigma}(0) - \underset{\sim}{\sigma}'_{p-1}\underset{\sim}{\phi}_{p-1}}.$$

The parameters $\sigma(0)$ and $\{\phi_p(p)\}$ are equivalent to the sequence $\{\sigma(h)\}$.

8. Prediction.

The autoregressive process (2) for $\mu = 0$ can be written

(77)
$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t,$$

where u_t is independent of $y_{t-1}, y_{t-2}, \ldots.$ Then the conditional expectation of y_t given $y_{t-1}, y_{t-2}, \ldots$ is

(78)
$$\mathcal{E}\{y_t | y_{t-1}, y_{t-2}, \ldots\} = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}.$$

This conditional expectation is the best predictor of y_t based on the past. Its variance is

(79)
$$\mathcal{E}[\phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - y_t]^2 = \mathcal{E}\, u_t^2 = \sigma^2.$$

The mean squared error of any other predictor, say $f(y_{t-1}, y_{t-2}, \ldots)$, is

(80)
$$\mathcal{E}\left[f(y_{t-1}, y_{t-2}, \ldots) - y_t\right]^2$$

$$= \mathcal{E}\left[f(y_{t-1}, y_{t-2}, \ldots) - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p} - u_t\right]^2$$

$$= \mathcal{E}\, u_t^2 + \mathcal{E}\left[f(y_{t-1}, y_{t-2}, \ldots) - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p}\right]^2.$$

Theorem 5. The predictor of y_t based on $y_{t-1}, y_{t-2}, \ldots$ with minimum mean square error for the stationary stochastic process satisfying (2) for which the roots are less than 1 in absolute value is (58).

References

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley and Sons, Inc., New York.

Technical Reports

U.S. Army Research Office - Contracts DAAG29-82-K-0156 and DAAG29-82-K-0239

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T. W. Anderson and Kai–Tai Fang, September 1982.

2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.

3. "Some Further Applications of Finite Difference Operators," Kai–Tai Fang, September 1982.

4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.

5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai–Tai Fang and Shu–Dong He, October 1982.

6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.

7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.

8. "Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

10. "Relationships Among Classes of Spherical Matrix Distributions," Kai–Tai Fang and Han–Feng Chen, April 1984.

11. "A Generalization of Autocorrelation and Partial Autocorrelation Functions Useful for Identification of ARMA(p,q) Processes," Akimichi Takemura, May 1984.

12. "Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models," T. W. Anderson and N. D. Singpurwalla, October 1984.

13. "Why Do Noninvertible Estimated Moving Averages Occur?" T. W. Anderson and Akimichi Takemura, November 1984.

14. "Invariant Tests and Likelihood Ratio Tests for Multivariate Elliptically Contoured Distributions," Huang Hsu, May 1985.

15. "Statistical Inferences in Cross–lagged Panel Studies," Lawrence S. Mayer, November 1985.

16. "Notes on the Extended Class of Stein Estimators," Suk–ki Hahn, July 1986.

17. "The Stationary Autoregressive Model," T. W. Anderson, July 1986.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER No. 17	**2. GOVT ACCESSION NO.** N/A	**3. RECIPIENT'S CATALOG NUMBER** N/A
4. TITLE (and Subtitle) The Stationary Autoregressive Model		**5. TYPE OF REPORT & PERIOD COVERED** Technical **6. PERFORMING ORG. REPORT NUMBER** --
7. AUTHOR(s) T. W. Anderson		**8. CONTRACT OR GRANT NUMBER(s)** DAAG29-85-K-0239
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall Stanford University Stanford, CA 94305		**10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS** P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		**12. REPORT DATE** July 1986 **13. NUMBER OF PAGES** 21 pp.
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		**15. SECURITY CLASS. (of this report)** Unclassified **15a. DECLASSIFICATION/DOWNGRADING SCHEDULE**

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Autoregressive model, stationary process, Yule-Walker equations, autocovariance sequence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(See reverse side for Abstract.)

20. Abstract.

The properties of the autoregressive process are obtained in terms of a stationary stochastic process satisfying the simple condition that a linear combination of current and past elements of the process is independent of all earlier elements of the process. The stationarity and independence imply that the roots of the associated polynomial equation are less than 1 in absolute value. The existence of the moving average representation is deduced, and the coefficients are derived for distinct roots. The Yule-Walker equations, which are derived, determine the autocovariance sequence. Another set of parameters consists of the variance of the process and the partial autocorrelation sequence.